Check for updates

Commentary: Conserved polymorphic sequences protect themselves for future challenges

Roger L. Dawkins* and Sally S. Lloyd

CY O'Connor ERADE Village Foundation, North Dandalup, WA, Australia

A Commentary on
Speciation of genes and genomes: Conservation of DNA polymorphism by barriers to recombination raised by mismatch repair system
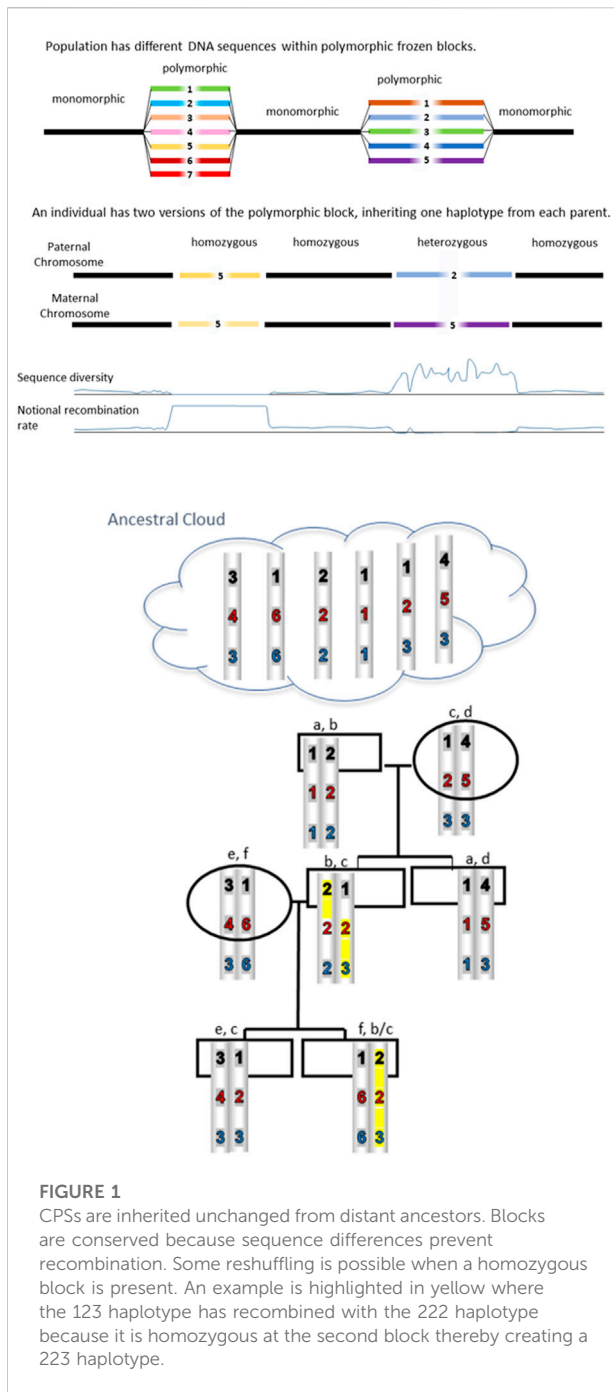
by Radman M (2022). Front. Genet. 13:803690. doi: 10.3389/fgene.2022.803690

Introduction

The Research Topic "Population genomic architecture: Conserved polymorphic sequences (CPSs), not linkage disequilibrium" addresses the discovery of distinct genomic regions or blocks which are highly polymorphic and therefore different between random individuals (Figure 1). Importantly, these differences are not due to recent mutation but have been conserved over hundreds of generations. So as avoid some of the confusion in the literature, we use mutation to describe an observed change in the sequence whereas polymorphism refers to a sequence difference which is inherited faithfully (Dawkins et al., 2013). This distinction is important but often overlooked. Polymorphisms which are "fixed" or "frozen" must somehow be protected from background mutation and recombination within the CPS. Just how this might happen has been of great interest and subject to much debate over several decades. For reviews see Steele (2014), Dawkins (2015), and Dawkins (2022).

Conserved polymorphic sequences

The initial discoveries of Conserved Extended Haplotypes are presented in the chapter by Alper (2021). By the 1970s it was known that there could be extensive HLA haplotypes containing specific alleles of HLA-A and HLA-B which are nearly 1 Mb apart but inherited together, without recombination. It was soon appreciated that such subjects

FIGURE 1
CPSs are inherited unchanged from distant ancestors. Blocks are conserved because sequence differences prevent recombination. Some reshuffling is possible when a homozygous block is present. An example is highlighted in yellow where the 123 haplotype has recombined with the 222 haplotype because it is homozygous at the second block thereby creating a 223 haplotype.

with HLA-A1 and HLA-B8 frequently had DR3 bringing the length of the A1, B8, DR3 or 8.1 haplotype to almost 2 Mb containing multiple genes, many of them duplicated. At the time there was much speculation as to how such "linkage disequilibrium" could be explained. Was there some form of *cis* interaction between the gene products? However, all became clear when the Alper group showed that individuals with 8.1 also have a specific complotype (SC01), unrelated to HLA, midway between HLA B and DR (Alper et al., 1982).

It was only a matter of time before methods, such as pulsed field electrophoresis (Tokunaga et al., 1989), allowed the direct demonstration of such an extensive DNA sequence. It transpired that there are many other ancestral haplotypes, such as 57.1 (A1, B57, DR3) which must have been conserved since before the Out-of-Africa migration i.e., during thousands of generations and meioses (Kay et al., 1988). There were other interesting examples such as 7.1 (A3, B7, DR15) which contrasted with 8.1 in that it is protective with respect to Insulin Dependent (or Juvenile) Diabetes Mellitus (Degli-Esposti et al., 1992). How could this be explained?

By the 1990s, it was obvious that these ancestral haplotypes did more functionally than simply carrying HLA class I and II alleles. For example, 8.1 affects the concentrations of TNF and IgA in the plasma and the development of hyperplasia within the thymus (French and Dawkins, 1990; Vandiedonck et al., 2004). So, there are still many fascinating questions to be resolved. How do these conserved sequences do what they do?

Some of these questions are being addressed by Okano et al. (2020), Kulski et al. (2021a), Kulski et al. (2021b), Cun et al. (2021), and Tay et al. (2021). These groups are discovering new mechanisms of gene regulation embedded within the non-coding regions of Ancestral Haplotypes. As proposed by Kulski, Retroviral like elements may have several roles. Firstly, they can generate sequences involved in *cis* and *trans* interactions. Secondly, they can introduce conserved polymorphism associated with duplication.

Mechanisms

Fortunately, we now have the benefit of Miro Radman's provocative proposal based on his pioneering work on the actual mechanisms involved in mutation and recombination (Radman, 2022). Perhaps polymorphism can be self-sustaining. "The mismatching of DNA sequences and the recognition of mismatched base pairs by mismatch repair (MMR) proteins are the determinants of …. conservation of highly polymorphic genetic motifs." The sequence differences and the activity of the MMR prevent recombination which would otherwise fragment the CPS.

How much sequence difference is required in order to conserve polymorphism by preventing recombination? Radman concludes that a minimal figure is of the order of 1 difference for every 30 base pairs in bacteria and 1/200 bp in lower eukaryotes. In mammals, the figure may be 1/400 bp. Interestingly, earlier comparisons of MHC ancestral haplotypes revealed sequence diversity sufficient to prevent recombination according to Radman's predictions (Gaudieri et al., 1997, 1999; Longman-Jacobsen et al., 2003; Lloyd et al., 2016). The careful study by Smith et al. (2006) was also consistent with the Radman number for mammals. In a landmark study (Traherne et al., 2006), cite 3.4 to 3.6 differences per 1,000 bp,

suggesting that there may be a buffer or safety zone between 1/300 and 1/400. Further, it appears that a linear increase in sequence divergence leads to exponential decrease in homologous recombination. The figure illustrates how Radman's predictions result in conservation of polymorphic haplotypes. Blocks are conserved because sequence differences prevent recombination; these haplotypes are inherited unchanged from distant ancestors. However, as shown in the figure, some reshuffling is possible when a homozygous block is present, with the implication that frequent CPS can diverge.

No doubt there are additional factors and species differences, but it does appear reasonable to conclude that a certain amount of polymorphism, *ipso facto*, results in conservation. Critical, however, is that this safety zone for conservation must be retained over generations.

Implications

The implications must be profound and extend to other genomic regions in Humans and, presumably, other vertebrates. We see these CPSs as the keys to an understanding of the fate of different individuals within the species. They can provide a type of memory for recurring challenges. They allow a concept which complements survival of the fittest with survival of the experienced.

Thus, the availability of tried and tested CPS permits some optimism, at least with respect to recurring events whether infectious or environmental and, also, with respect to changes in food and behavior. There is less cause for optimism in the case of new man-made challenges, such as population growth and biological warfare, as just two examples.

The further implications of CPSs deserve attention and might include, *inter alia*:

(a) Heterozygous advantage.
(b) Shuffling within the monomorphic sequence (see Figure 1) so as to increase diversity.
(c) Caveats on the use and interpretation of studies of homozygous cell lines.
(d) Strategies for threatened populations.

(e) Selection of preferred traits in livestock.
(f) Self vs. non-self recognition and autoimmunity.

Conclusion

Conservation of Polymorphic Sequences associated with gene duplication may achieve the Radman number and thereafter provide stability of inheritance of multigenic traits.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alper, C. A., Awdeh, Z. L., Raum, D., and Yunis, E. J. (1982). Extended major histocompatibility complex haplotypes in man: Role of alleles analogous to murine t mutants. *Clin. Immunol. Immunopathol.* 24, 276–285. doi:10.1016/0090-1229(82)90238-0

Alper, C. A. (2021). The path to conserved extended haplotypes: Megabase-length haplotypes at high population frequency. *Front. Genet.* 12, 716603–716608. doi:10.3389/fgene.2021.716603

Cun, Y., Shi, L., Kulski, J. K., Liu, S., Yang, J., Tao, Y., et al. (2021). Haplotypic associations and differentiation of MHC class II polymorphic alu insertions at five loci with HLA-DRB1 alleles in 12 minority ethnic populations in China. *Front. Genet.* 12, 636236. doi:10.3389/fgene.2021.636236

Dawkins, R. L. (2015). *Adapting Genetics: Quantal evolution after natural selection - surviving the changes to come.* Dallas, TX: Near Urban Publishing.

Dawkins, R. L. (2022). *Adapting genetics. Educating the genome: Safe deposit evolution.* Second. Dallas, TX: Near Urban LLC.

Dawkins, R. L., Willamson, J. F., Lester, S., and Dawkins, S. T. (2013). Mutation versus polymorphism in evolution. *Genomics* 101, 211–212. doi:10.1016/j.ygeno.2013.01.002

Degli-Esposti, M. A., Abraham, L. J., Mccann, V., Spies, T., Christiansen, F. T., and Dawkins, R. L. (1992). Ancestral haplotypes reveal the role of the central MHC in the immunogenetics of IDDM. *Immunogenetics* 36, 345–356. doi:10.1007/BF00218041

French, M. A., and Dawkins, R. L. (1990). Central MHC genes, IgA deficiency and autoimmune disease. *Immunol. Today* 11, 271–274. doi:10.1016/0167-5699(90)90110-u

Gaudieri, S., Kulski, J. K., Dawkins, R. L., and Gojobori, T. (1999). Extensive nucleotide variability within a 370 kb sequence from the central region of the Major Histocompatibility Complex. *Gene* 238, 157–161. doi:10.1016/S0378-1119(99)00255-3

Gaudieri, S., Leelayuwat, C., Tay, G. K., Townend, D. C., and Dawkins, R. L. (1997). The major histocompatability complex (MHC) contains conserved polymorphic genomic sequences that are shuffled by recombination to form ethnic-specific haplotypes. *J. Mol. Evol.* 45, 17–23. doi:10.1007/PL00006194

Kay, P. H., Dawkins, R. L., Williamson, J., Tokunaga, K., Christiansen, F. T., and Charoenwong, P. (1988). Coexistence of an MHC chromosomal segment marked by HLA B17, BfS, C4A6, B1, DR7, and DQw9 in different ethnic groups. *Hum. Immunol.* 23, 27–35. doi:10.1016/0198-8859(88)90015-8

Kulski, J. K., Suzuki, S., and Shiina, T. (2021a). Haplotype shuffling and dimorphic transposable Elements in the human extended major histocompatibility complex class II region. *Front. Genet.* 12, 1–22. doi:10.3389/fgene.2021.665899

Kulski, J. K., Suzuki, S., and Shiina, T. (2021b). SNP-density crossover maps of polymorphic transposable Elements and HLA genes within MHC class I haplotype blocks and junction. *Front. Genet.* 11, 594318. doi:10.3389/fgene.2020.594318

Lloyd, S. S., Steele, E. J., and Dawkins, R. L. (2016). "Analysis of haplotype sequences," in *Next generation sequencing - advances, applications and challenges*. Editor J. K. Kulski (London: InTech), 345–368. doi:10.5772/60489

Longman-Jacobsen, N., Williamson, J. F., Dawkins, R. L., and Gaudieri, S. (2003). In polymorphic genomic regions indels cluster with nucleotide polymorphism: Quantum genomics. *Gene* 312, 257–261. doi:10.1016/S0378-1119(03)00621-8

Okano, M., Miyamae, J., Suzuki, S., Nishiya, K., Katakura, F., Kulski, J. K., et al. (2020). Identification of novel alleles and structural haplotypes of major histocompatibility complex class I and DRB genes in domestic cat (*Felis catus*) by a newly developed NGS-based genotyping method. *Front. Genet.* 11, 750–815. doi:10.3389/fgene.2020.00750

Radman, M. (2022). Speciation of genes and genomes: Conservation of DNA polymorphism by barriers to recombination raised by mismatch repair system. *Front. Genet.* 13, 803690–803712. doi:10.3389/fgene.2022.803690

Smith, W. P., Vu, Q., Li, S. S., Hansen, J. a., Zhao, L. P., and Geraghty, D. E. (2006). Toward understanding MHC disease associations: Partial resequencing of 46 distinct HLA haplotypes. *Genomics* 87, 561–571. doi:10.1016/j.ygeno.2005.11.020

Steele, E. J. (2014). Reflections on ancestral haplotypes: Medical genomics, evolution, and human individuality. *Perspect. Biol. Med.* 57, 179–197. doi:10.1353/pbm.2014.0014

Tay, G. K., Al Naqbi, H., Mawart, A., Baalfaqih, Z., Almaazmi, A., Deeb, A., et al. (2021). Segregation analysis of genotyped and family-phased, long range MHC classical class I and class II haplotypes in 5 families with type 1 Diabetes proband in the united Arab emirates. *Front. Genet.* 12, 670844. doi:10.3389/fgene.2021.670844

Tokunaga, K., Kay, P. H., Christiansen, F. T., Saueracker, G., and Dawkins, R. L. (1989). Comparative mapping of the human major histocompatibility complex in different racial groups by pulsed field gel electrophoresis. *Hum. Immunol.* 26, 99–106. doi:10.1016/0198-8859(89)90095-5

Traherne, J. A., Horton, R., Roberts, A. N., Miretti, M. M., Hurles, M. E., Stewart, C. A., et al. (2006). Genetic analysis of completely sequenced disease-associated MHC haplotypes identifies shuffling of segments in recent human history. *PLoS Genet.* 2, e9–e92. doi:10.1371/journal.pgen.0020009

Vandiedonck, C., Beaurain, G., Giraud, M., Hue-Beauvais, C., Eymard, B., Tranchant, C., et al. (2004). Pleiotropic effects of the 8.1 HLA haplotype in patients with autoimmune myasthenia gravis and thymus hyperplasia. *Proc. Natl. Acad. Sci. U. S. A.* 101, 15464–15469. doi:10.1073/pnas.0406756101